DOE Next Generation Internet Program

DOE QoS Application Testbed Status

(http://www.itg.lbl.gov/NGI/status
and
http://www.es.net/publish/ngi-bb.html)

William Johnston, LBNL
Rebecca Nitzan, LBNL/ESNet
Linda Winkler, ANL

Goals

- Explore potential new applications when network QoS and bandwidth reservation are routinely available
- Provide an application requirements driven, QoS technology development environment
- Explore the issues of building a nation-wide, multi-domain QoS testbed

Approach

- Direct involvement of the application community
- QoS architecture
  - reservation based IP differentiated services
  - resource managers providing access to router QoS functionality
  - policy based access control for “premium” service
  - application interfaces to bandwidth reservation based on established Globus services
  - should also work with dynamic ATM SVC, etc.
- A testbed architecture and implementation:
  - use existing ESNet circuits
  - place end-nodes in the application environment

Applications

- Corridor One: An Integrated Distance Visualization Environment for SSI and ASCI Applications
  - low latency for collaboration interaction, and immersive applications
- Prototyping an Earth System Grid: Enable a geographically distributed community to perform sophisticated, computationally intensive analyses of Petabytes of climate data.
  - 100 Mbit/sec bandwidth reservation for distributed “interactive” model analysis and visualization.

Applications (cont.)

- **Grid-based X-Ray Crystallography Collaboratory: Instrument data stream analysis, remote storage, and feedback control**
  - **128Mbit/sec bandwidth reservation**
- **Prototyping a Combustion Corridor: Real time interactive volume visualization of combustion data sets**
  - **16 Mbit/sec bandwidth reservation, low latency**

Applications (cont.)

- **EMERGE: ESnet/MREN Regional Grid Experimental NGI Testbed: Designing, deploying and testing differentiated services on an IP/ATM regional GigaPoP network interoperating with ESnet for applications in combustion, climate and high-energy physics**
  - **10-30 Mbit/sec bandwidth reservation, low latency**
- **Particle Physics Data Grid: High-speed WAN-based storage and retrieval of high energy physics data**
  - **guaranteed delivery of large volumes of data within a specified period of time (average bandwidth requirements are large - O(800 Mbit/s)**

Issues Being Addressed

- Deploying the QoS testbed throughout ESNet and connecting to Abilen and MREN
- Bandwidth must be reserved throughout a network domain with a global view of all commitments within that domain
  - to avoid overloading internal network nodes with premium traffic
  - provide end-to-end application guarantees

Issues (cont.)

- Inter ISP "brokering"
  - bandwidth reservations are only useful to applications if that are end-to-end
  - Interoperating with I2 QBone is the focus for this
- What policy should be used to allocate premium bandwidth?
  - (more later on this)
- What is necessary to make the reservation and claiming processes scalable?
  - (more on this later)

Issues (cont.)

- **All access points where premium bandwidth could be available must have access control**
  - **protect against mis-appropriation of premium bandwidth**
- **Traffic shaping elements will be necessary**
  - **end hosts can traffic shape only at a coarse level (e.g. clocking socket writes)**
  - **to avoid triggering the policer that protects the IPS from too much premium traffic these streams have to be shaped at the IP packet (maybe ATM cell) level before entering the QoS network**

- **Current ESNet allocation of bandwidth to the testbed is via UBR: This lets the testbed use up to X Mbit/s of the production circuits, but does not guarantee that the production traffic will not intrude on the testbed bandwidth**
  - **links will be monitored during QoS experiments to see if production traffic has interfered with the QoS bandwidth (and thus invalidated the experiment)**
  - **this not anticipated to be a problem in the near term as most of the production links are at less than capacity on the ATM circuits**
  - **persistent and guaranteed QoS bandwidth will be needed in future**

QoS Architecture

- **Site Resource Manager**
  (Little or no real experience with this function and several different approaches are being explored.)
  - **accepts reservation requests from applications**
  - **locates and/or obtains commitments from the required resources within the domain**
  - **provides the application flow spec to the shaper and ISP ingress router when the reserved bandwidth traffic turns on ("claiming")**

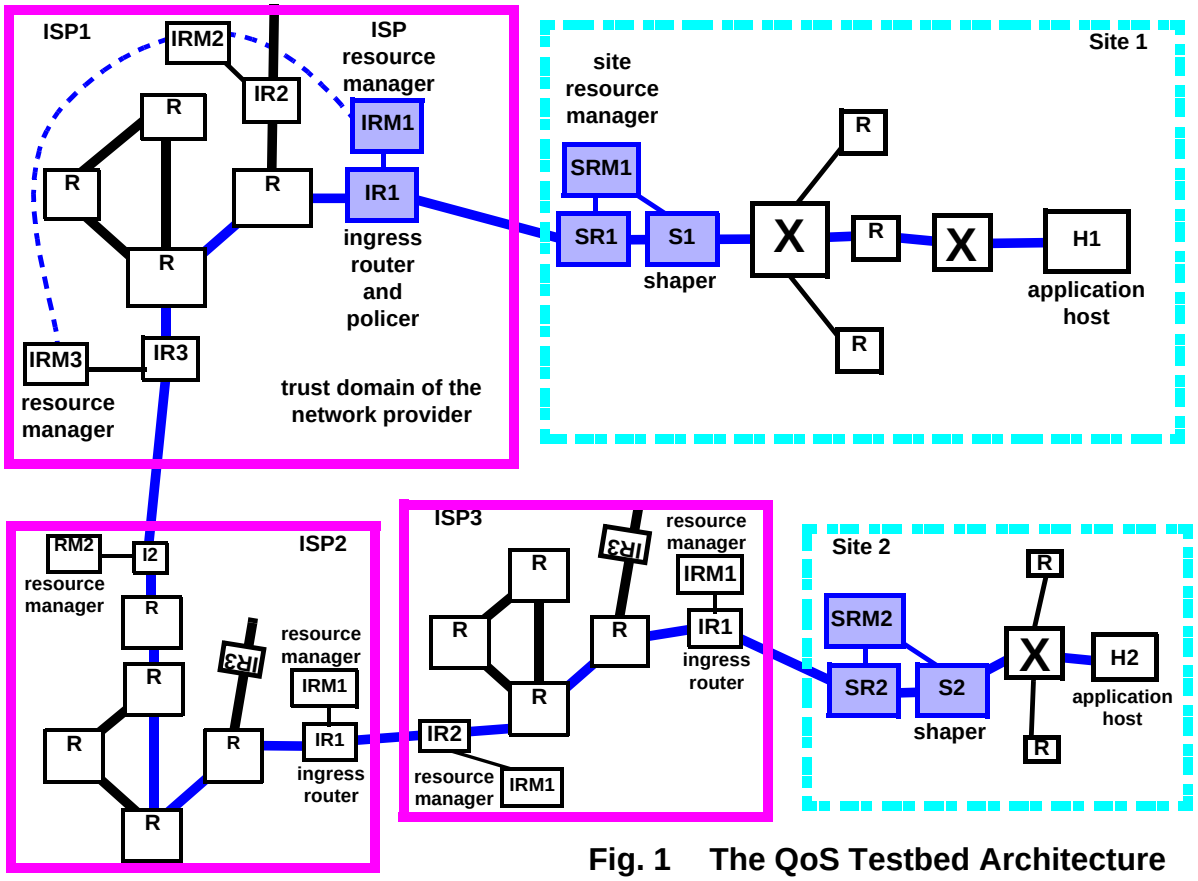


Fig. 1 The QoS Testbed Architecture

QoS Architecture (cont.)

- **Shaper**
  - **takes application traffic that is within the agreed on limit at the level of, e.g., socket writes and produces a uniform stream of IP packets or ATM cells, as required**
- **ISP Resource Manager**
  - **"commits resources for a reservation" in domain (again, several approaches being investigated)**
- **ISP Ingress Router**
  - **polices premium bandwidth streams to be sure that they are within the reservation specs**

An interesting (perhaps central) allocation management policy issue for ISP like ESNet

- **How do you manage the allocation of available premium bandwidth among your customers? (can probably assume that the customers are not going to set up a global allocation committee - guaranteed not to happen once more than one ISP is involved)**
- **Assume that an ISP (e.g. ESNet) has configured to support 50 Mb/s premium traffic internally**

Allocation Policy (cont.)

- **Assume that we don't want to address this issue by taking all of the ISP customers and allocating each 1/N X 50 Mb/s - you want to accommodate "large-scale" applications, at least when others are not using their allocation**
- **On the other hand, you don't want to allocate all of the premium bandwidth on a first come, first served basis**
- **Compromise: guarantee each customer a certain minimum premium bandwidth, and allow everything above this to be reserved on a first come basis for large apps.**

Allocation Policy (cont.)

- **What does it take to implement this?**
  - **an allocation policy enforced at reservation time (Requires an ISP to have a global view of allocated resources within its domain: Bandwidth available for reservation changes dynamically.)**
  - **the policing function must also change its target bandwidth dynamically so that the sum is always fixed (assuming a homogeneous network interior)**

What is necessary to make the reservation and claiming processes scalable?

- **The reservation process must be automated – this will probably involve brokers and policy based access control**
- **Different domains will have different policies and different consumer populations**
- **Even in simplest case (e.g., two ESNet sites) there are likely to be three policies involved:**
  - **site_1's control of its allocation**
  - **ESNet's control of its allocation strategy**
  - **site_2's control of its allocation**

Scalable Reservation (cont.)

- **At each domain boundary the appropriate member of that domain's consumer population must make the request in order to satisfy the domain policy**
- **One way to address this for end user reservation requests is with "delegated authorization"**
- **Management of the delegated authorizations by a broker can then automate the process**

For example, user_1 at site_1 and user_2 at site_2 wish to communicate using premium bandwidth:

Scalable Reservation (cont.)

- user_1 engages a broker to make end-to-end and bi-directional reservation
- user_1 satisfies site_1 policy and a reservation is made with site_1 resource manager
- site_1 resource manager delegates it authorization with ESNet to user_1 (user_1 is not a customer of ESNet, site_1 is)
- broker contacts ESNet ingress resource manager which validates the request based on its policy that applies to site_1
- similarly user_2 delegates it's site_2 authorization to user_1, which is used to obtain the reverse path reservation from site_2

Testbed Implementation Status

¨ See http://www.es.net/publish/ngi-bb.html

¨ Routers

- Weighted Fair Queuing (the basic QoS mechanism) works and is deployed at four sites
- congestion experiments do not always produce good results, however things seems to basically work at < 40 Mbits/s
- router QoS management interfaces are unstable
- NGI QoS testbed routers are being deployed as in table, below

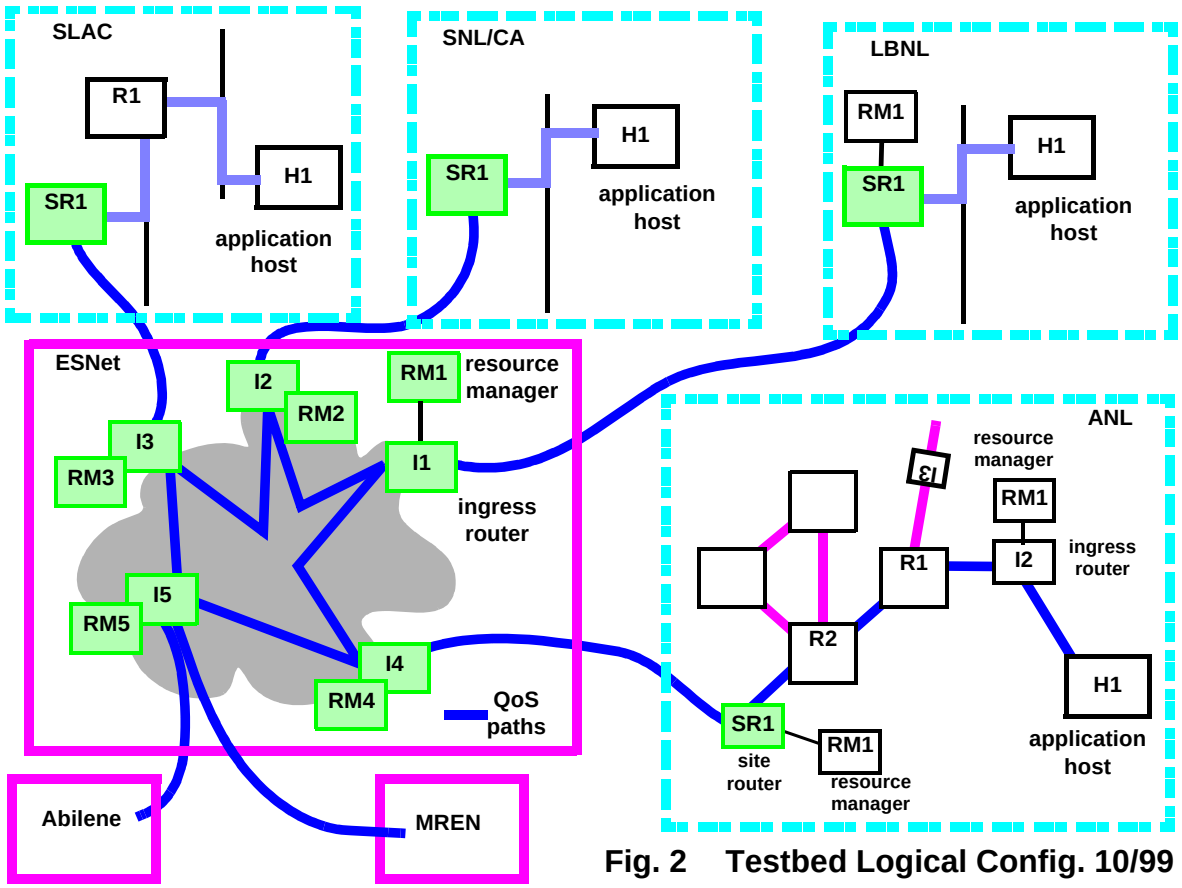
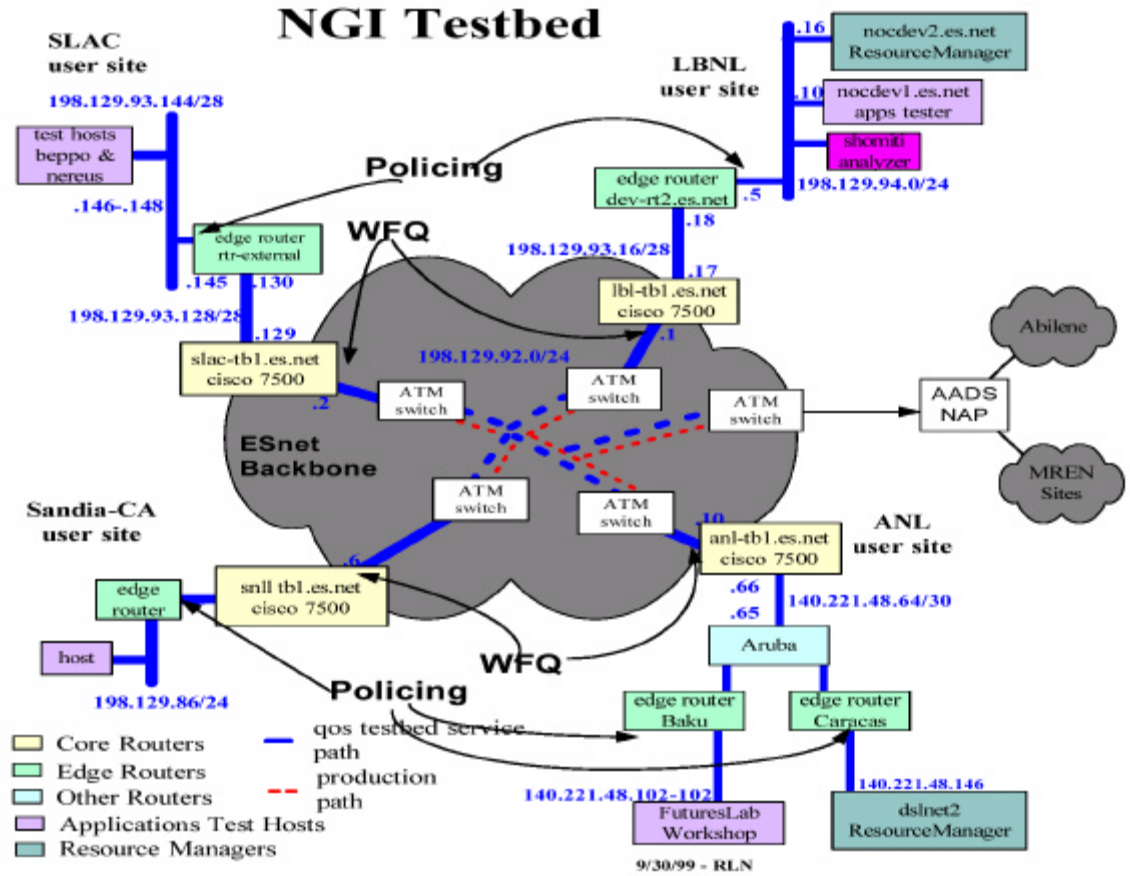
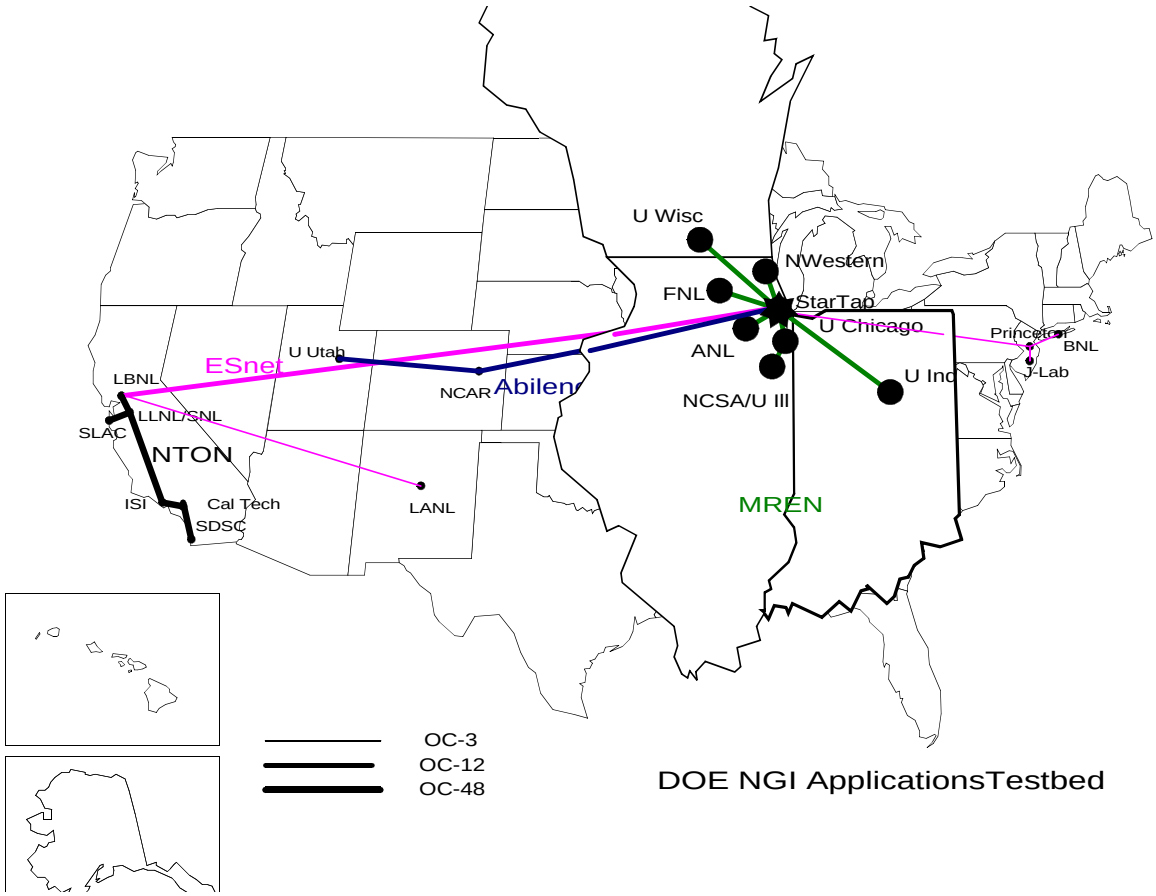


Fig. 2 Testbed Logical Config. 10/99





		Connected?	when?
ANL	ESnet	YES	DONE
BNL	ESnet	No	Est 11/1/99
FNAL	ESnet	No	Est 11/1/99
Indiana University - Indianapolis	Abilene	No	Est 11/1/99
Indiana University - Bloomington	Abilene	No	Est 3/1/00
JLAB	ESnet	No	Est 11/1/99
LANL	ESnet	No	Est 11/1/99
LBNL	ESnet	YES	DONE
NCAR	Abilene	No	Est ?
Northwestern	MREN	No	Est 11/1/99
Princeton	vBNS	No	Est ?
Sandia - Livermore	ESnet	Yes	DONE
SDSC	Calren2, NTON	No	Est ?
SLAC	ESnet, NTON	Yes	DONE
University of Illinois at Chicago	MREN	No	Est 11/1/99
University of Illinois at Urbana Champaign	MREN	No	Est 11/1/99
University of Utah	Abilene	No	Est ?
University of Chicago	MREN	No	Est 11/1/99
University of Wisconsin	MREN	No	Est 11/1/99
USC/ISI	Calren2?	No	Est ?

DOE NGI Site Status